Centres for European Policy Network
FREIBURG | BERLIN | PARIS | ROMA

cepInput

No 7 | 2024 16 aprile2024

Le tecnologie dei modelli linguistici, un vantaggio competitivo per l'UE

Dieci punti per un uso ottimale dell'IA generativa nelle PMI

Anselm Küsters

I modelli linguistici come ChatGPT rappresentano una grande sfida, ma anche un'opportunità per l'Europa. Invece di rispondere con il protezionismo, progetti faro o l'avversione al rischio è necessario un approccio pragmatico all'uso su larga scala delle tecnologie dei modelli linguistici di intelligenza artificiale nell'economia, al fine di mantenere la competitività e sfruttare il potenziale di innovazione. Questo input del Cep delinea dieci fattori che le piccole e medie imprese (PMI) dovrebbero considerare per sfruttare i vantaggi competitivi esistenti.

- Le PMI dovrebbero capire concettualmente come l'IA possa migliorare i loro processi attraverso l'analisi dei bisogni e la pianificazione strategica. La scelta di servizi basati su cloud e l'uso di modelli aperti influenzano l'adattabilità degli strumenti di IA e la conseguente dipendenza da aziende esterne.
- Grazie al "*fine tuning*" e all'uso della "generazione aumentata di recupero" (RAG), le PMI possono specializzare le loro applicazioni. Il rischio di errore dei modelli di IA deve essere allineato alla tolleranza interna all'errore. Inoltre, è necessario sviluppare competenze interne di "*prompt design*" e "*online design*".
- Per un uso sostenibile e socialmente responsabile dell'IA, occorre tenere conto del quadro normativo, effettuare continui test interni e misurare l'efficienza energetica. Inoltre, è necessario utilizzare meccanismi appropriati per mantenere la tecnologia aggiornata.

Indice

1 Introduzione: utilizzare le tecnologie linguistiche invece di riprodurle

Per ogni domanda inviata a ChatGPT viene utilizzato un modello linguistico di grandi dimensioni. Di fronte allo sviluppo vertiginoso del settore dell'intelligenza artificiale (IA), l'Europa si trova di fronte a una sfida importante, che finora è stata affrontata ponendo troppa enfasi sui progetti faro[1] e non abbastanza sulla sua applicazione su larga scala nell'economia. Attualmente, la politica europea si concentra sull'obiettivo di assicurarsi una quota quanto più ampia possibile della catena del valore dell'IA nel lungo periodo, per evitare una dipendenza strategica in una fase successiva. Ciò include la costruzione di adeguato spazio dati[2], la sovvenzione massiccia di fabbriche di chip[3] e, più recentemente, la gestione di supercomputer di IA, la costruzione di cosiddette *"fabbriche di IA* "[4] e una "Alleanza europea per le tecnologie linguistiche" per costruire i propri modelli linguistici principali[5]. La speranza, a malapena celata, di creare i propri campioni nazionali ha guidato Paesi come la Germania e la Francia nei negoziati finali sulla legge europea sull'IA[6]. Nel complesso, queste iniziative per costruire le proprie tecnologie linguistiche, devono essere intese come parte di una tendenza più ampia di *"home-shoring"*[7] che sta diventando sempre più importante nel contesto delle tensioni geopolitiche dall'Ucraina a Taiwan.

Anche se una simile riflessione strategica - a lungo trascurata a livello europeo è utile nel lungo periodo, le dinamiche dell'IA non ci permettono di rimandarne ulteriormente l'applicazione pratica. Lo sviluppo esponenziale della tecnologia, finora concentrato negli Stati Uniti[8], ci costringe a trovare "soluzioni di ripiego" diverse dall'approccio politico adottato finora in Europa. Invece di mettere in atto esclusivamente piani per lo sviluppo di infrastrutture e modelli, accompagnati da lente procedure di appalto pubblico e da una dettagliata regolamentazione europea, dovremmo ora concentrarci maggiormente dell'attuazione pratica di questa tecnologia. Data la difficile situazione economica e la battaglia per la competitività globale, **le piccole e medie imprese (PMI) europee in particolare non possono aspettare che i fornitori nazionali sviluppino modelli competitivi**. I modelli commerciali statunitensi e i modelli *open source* gratuiti sono preziosi per un'applicazione rapida e diffusa delle tecnologie linguistiche dell'IA.

Sebbene l'IA generativa abbia il potenziale per creare dai 2,6 ai 4,4 trilioni di dollari di valore aggiunto in tutti i settori, grazie ai suoi numerosi casi di applicazione[9], l'integrazione dei modelli esistenti nella pratica commerciale è stata finora carente ad esempio in Paesi come la Germania.[10] Anche se si stima che strumenti di IA come ChatGPT potrebbero automatizzare il 60-70% del tempo di lavoro dei “colletti bianchi”, i manager sono stati finora piuttosto reticenti, anche perché la tecnologia non è

[1] Su questa critica ai "fari", si veda: Friesike e Sprondel (2022), Träg Transformation. Welche Denkfehler den digitalen Wandel blockieren, Stoccarda: Reclam

[2] European Common Data Spaces | Shaping Europe's digital future (europa.eu).

[3] Küsters e Kullas (2023), Can the Chips Act make Europe more resilient, Audit Committee Quarterly II/2023.

[4] La Commissione lancia il pacchetto innovazione AI (europa.eu).

[5] Studio di fattibilità LEAM 2023 - KI-Verband; Launching an 'AI moonshot' to develop a European large language model is the game changer that Europe needs - CEPS.

[6] Regole UE per ChatGPT e Aleph Alpha: Germania e Francia contrarie (faz.net).

[7] Foroohar (2022), Homecoming: The Path to Prosperity in a Post-Global World, Penguin Random House.

[8] The Economist (2024), How San Francisco staged a surprising comeback (12 febbraio 2024).

[9] Si vedano le statistiche su: Potentiel économique de l'AI générative | McKinsey.

[10] Si veda ad esempio: Un sondaggio mostra che l'IA è in una cattiva posizione nelle aziende tedesche - Tagesspiegel Retroscena. In generale, sul ritardo: Gutachten zu Forschung, Innovation und Technologischer Leistungsfähigkeit Deutschlands 2024 (e-fi.de), pag. 116 e seguenti.

considerata matura o è ancora troppo imprecisa[11]. Nonostante le preoccupazioni per gli errori e i rischi addirittura esistenziali associati all'IA, vi è il pericolo di non cogliere i benefici delle tecnologie dell'IA a causa di un'eccessiva cautela, come hanno recentemente sottolineato anche le Nazioni Unite.[12] In effetti, i progressi metodologici che sono stati compiuti in un arco di tempo molto breve nel campo dell'IA generativa sono spesso trascurati. Il timore, anch'esso spesso espresso, di cadere in una successiva dipendenza nell'integrazione di tecnologie straniere sembra al momento meno urgente, data la forte pressione competitiva. Inoltre, questo rischio strategico è meno significativo di quanto si pensi. La disponibilità di moderni modelli linguistici consente a piccoli team non specializzati di costruire applicazioni flessibili, senza bisogno di codici o moduli specifici[13]. Questo riduce notevolmente i costi di apprendimento, riduce il potenziale di dipendenza e offre anche una risposta alle sfide ambientali del mercato digitale[14].

È quindi giunto il momento per le PMI europee di integrare modelli linguistici IA di alta qualità nei loro processi interni ed esterni. Questo contributo del Cep vuole servire come panoramica concettuale dei fattori da considerare e delinea una strategia di tecnologia linguistica per le PMI in 10 punti, che può essere utilizzata come base per lo sviluppo di una politica interna sull'uso di strumenti di IA generativa.[15] Questi elementi vanno dalla progettazione delle interrogazioni ("prompt") alle preoccupazioni relative alla protezione dei dati. In generale, questa pubblicazione non intende fornire consulenza legale, ma fornire informazioni sulle possibilità e sulle applicazioni di questi nuovi strumenti ed i relativi problemi. Naturalmente, l'utilità o meno dell'uso dei modelli linguistici dal punto di vista aziendale dovrà essere valutata caso per caso. Tuttavia queste "tecnologie di uso generale"sono già utilizzate su larga scala e in tutti i settori e, secondo le attuali stime della letteratura, influenzeranno direttamenteil 10-30% di tutti i lavoratori in Europa[16]. In questo contesto, l'integrazione delle tecnologie linguistiche non solo offre l'opportunità di aumentare l'efficienza e la capacità di innovazione, ma anche di rafforzare l'Europa nella competizione globale. Le PMI dovrebbero agire in modo proattivo per cogliere i vantaggi della tecnologia AI, valutandone attentamente i rischi.

2 Fattori di utilizzo dell'IA generativa nelle PMI

Che cos'è l'IA generativa e che cosa significano i modelli linguistici? Nel contesto di ChatGPT e simili, l'abbreviazione IA è oggi più spesso utilizzata per indicare una classe di modelli avanzati che vengono addestrati a produrre contenuti indistinguibili dal lavoro umano, siano essi testi, immagini, codici o persino brevi video. Il cuore di questa tecnologia sono i *Large Language Models* (LLM)[17] che, elaborando grandi quantità di testo, imparano a cogliere le sfumature del linguaggio umano e a

[11]Vedi le statistiche su: Top 30 Must Know Generative AI Stats in 2024 (aimultiple.com).
[12]UN AI Advisory Body, Interim Report: Governing AI for Humanity, December 2023, interim_report.pdf (un.org), p. 12.
[13]Si vedano anche le argomentazioni presentate di seguito nella sezione 2.1.
[14]Commissione (2024), Comunicazione della Commissione sulla definizione del mercato rilevante ai fini del diritto della concorrenza dell'UE, Bruxelles, 8.2.2024, C(2023) 6789 definitivo, punto 98.
[15]Questo non include la consulenza legale, ad esempio su questioni di protezione dei dati in sospeso. Per un esempio di politica interna sull'intelligenza artificiale, si veda ad esempio: BBC (2024), Guidance: The use of Artificial Intelligence (bbc.co.uk).
[16]Mauro Cazzaniga et al. (2024), Gen-AI : Artificial Intelligence and the Future of Work (imf.org). Albanesi, Stefania et Dias da Silva, Antonio et Jimeno, Juan F. et Lamo, Ana et Wabitsch, Alena, Nouvelles technologies et emplois en Europe (2023). Document de travail du NBER n° w31357.
[17]Come GPT ("Generative Pre-trained Transformer"). Per una panoramica, si veda: [2402.06196] Large Language Models: A Survey (arxiv.org).

utilizzarle per l'esplorazione creativa. I modelli linguistici definiscono una distribuzione di probabilità per le sequenze di parole e possono quindi essere utilizzati per scopi generativi, prevedendo le parole successive più probabili all'inizio di un testo. Negli ultimi anni, questi modelli sono diventati sempre più grandi (cioè si basano su un maggior numero di dati di addestramento e utilizzano più parametri nel modello), il che ha migliorato significativamente le loro capacità di predizione del testo (*"legge di scala* ")[18]. Questi modelli linguistici possono acquisire nuove capacità, come il calcolo, la risposta alle domande e la sintesi del testo attraverso l'osservazione del linguaggio naturale ("capacità emergenti")[19]. Gli enormi e imprevedibili balzi delle capacità di questi modelli hanno recentemente portato a un vero e proprio "boom dell'IA".

Grazie ai progressi descritti, i modelli più recenti possono non solo rendere le informazioni esistenti in modo coerente e contestuale, ma anche generare contenuti originali sulla base dei modelli appresi. Questo apre una vasta gamma di possibili applicazioni per le aziende, dalla scrittura automatica di testi alla generazione di opere creative e allo sviluppo di chatbot. Nel frattempo, gli esperti hanno messo insieme più di un centinaio di casi d'uso generali e specifici per il settore dell'IA generativa, molti dei quali potrebbero essere rilevanti anche per le PMI[20]. I processi che comportano un intenso lavoro con parole, immagini, numeri e suoni (il cosiddetto lavoro WINS, acronimo di *Words, Images, Numbers, Sounds*) sono quelli che dovrebbero beneficiare maggiormente della nuova tecnologia[21]. Gli strumenti di intelligenza artificiale generativa possono, ad esempio, facilitare la stesura di testi di marketing e di vendita, supportare lo sviluppo di idee di marketing creative, creare automaticamente modelli di documenti o riconoscere gli aggiornamenti normativi[22]. Un esempio particolarmente impressionante è fornito dal fornitore di servizi di pagamento Klarna, il cui assistente IA basato su ChatGPT si è fatto carico del lavoro di 700 persone a tempo pieno in un mese, risolvendo i problemi nelle discussioni con i clienti in modo più accurato, il che ha portato a un calo del 25% delle richieste di informazioni[23].

L'analisi che segue non si ferma a una valutazione generale dei vantaggi e degli svantaggi dell'IA generativa, ma cerca di individuare gli elementi chiave per una rapida integrazione dei modelli linguistici nei processi delle PMI europee e tedesche. Una strategia efficace per le tecnologie linguistiche dovrebbe considerare i seguenti dieci elementi.

2.1 Analisi dei bisogni: comprendere l'IA generativa da un punto di vista concettuale

Le PMI dovrebbero innanzitutto condurre un'analisi approfondita dei bisogni per capire quali processi interni ed esterni possono essere migliorati integrando i modelli linguistici. Attualmente, molte aziende non riescono a integrare i modelli linguistici perché non capiscono, dal punto di vista metodologico, che l'IA generativa non è una forma tradizionale di automazione, ma un agente di supporto che diventa più intelligente nel tempo.[24] Lo sviluppo dell'apprendimento automatico (ML) e la sua graduale diffusione nelle aziende nell'ultimo decennio offrono un interessante parallelo,

[18]Sardana et al. (2023), [2401.00448] Beyond Chinchilla-Optimal: Accounting for Inference in Language Model Scaling Laws (arxiv.org).
[19]Wei et al (2022), [2206.07682] Emergent Abilities of Large Language Models (arxiv.org).
[20]Top 100+ applicazioni e casi d'uso di IA generativa nel 2024 (aimultiple.com).
[21]Da dove dovrebbe iniziare la vostra azienda con l'IA generativa (hbr.org).
[22]Per i seguenti esempi, vedere : Potenziale economico dell'IA generativa | McKinsey.
[23]L'assistente AI Klarna gestisce due terzi delle conversazioni con il servizio clienti nel primo mese.
[24]La vostra organizzazione non è progettata per lavorare con GenAI (hbr.org).

mostrando quanto possa essere difficile passare dalla semplice fascinazione per una nuova tecnologia alla comprensione strategica delle sue applicazioni pratiche[25]. Sebbene le funzioni avanzate di ML, come il riconoscimento delle immagini e il riconoscimento vocale, siano diventate sempre più note negli anni 2010, molte aziende inizialmente non erano sicure di come utilizzare tali tecniche, soprattutto se non erano direttamente collegate al loro *core business*. Solo con il tempo la percezione del ML come strumento per compiti specifici è cambiata fino a diventare un sistema di riconoscimento dei modelli altamente sofisticato. Laddove i problemi esistenti potevano essere riformulati in problemi di riconoscimento dei modelli, le aziende e le *start-up* hanno creato nuove opportunità di business. Il valore aggiunto delle nuove tecnologie non sta quindi solo nell'integrazione nei processi aziendali esistenti, ma anche nella riprogettazione di tali processi.

L'IA generativa, come l'uso di modelli linguistici di grandi dimensioni, richiede ora un cambiamento concettuale simile alla graduale integrazione dell'IA. A quali problemi fondamentali in un determinato campo di attività può essere applicata questa tecnologia? A differenza dell'automazione tradizionale tramite robot, la funzionalità dell'IA generativa è meglio compresa attraverso la sua funzione di dialogo, che consente alla tecnologia e all'uomo di condividere le responsabilità in modo dinamico[26]. A tal fine, è utile immaginare l'IA generativa come l'equivalente di milioni di stagisti, cioè fantasiosi ed energici, imprecisi e in qualche modo imprevedibili, ma meno costosi e molto più scalabili degli stagisti reali[27]. Ad esempio, la letteratura sul *design thinking* finalizzata alla promozione dell'intelligenza istituzionale ha dimostrato che i risultati dei modelli linguistici di grandi dimensioni (LLM) dovrebbero essere considerati come idee e non come risposte definitive, e che questi sistemi dovrebbero quindi essere posizionati internamente come uno strumento di aiuto alla percezione umana[28]. In altre parole: mentre negli ultimi dieci anni l'apprendimento automatico classico ha potuto essere integrato nei processi aziendali grazie alla sua capacità di riconoscimento dei modelli, l'IA generativa sarà a disposizione delle aziende come partner di dialogo iterativo, paragonabile a un pool di apprendisti. Per le PMI, la prima questione concettuale che si pone è quindi come trasformarsi internamente per poter sfruttare al meglio questa funzione di dialogo. In questo senso, il valore del LLM risiede nella sua integrazione all'interno di sistemi più ampi e non nel suo utilizzo individuale.

La definizione degli obiettivi deve basarsi su questa analisi e stabilire traguardi chiari e misurabili per l'implementazione delle tecnologie linguistiche. Un buon esempio è il settore della pubblicità e del marketing, dove l'IA generativa viene già utilizzata per molte applicazioni, come la creazione di contenuti scritti e copy pubblicitari (58%), la ricerca di parole chiavi (43%) e i riassunti di e-mail, riunioni e campagne (38%)[29]. Un esempio impressionante è quello di Coca-Cola, che ha recentemente riferito di aver utilizzato l'IA generativa per creare automaticamente migliaia di contenuti di marketing. L'azienda ha deliberatamente spostato la sua spesa mediatica dagli annunci televisivi, che spesso richiedevano mesi per essere prodotti e non potevano essere modificati in seguito, ai canali digitali per i quali la tecnologia dei modelli linguistici ha prodotto circa "1.000 contenuti contestualmente rilevanti", i cui risultati erano anche misurabili in tempo reale[30]. L'impatto di questa spesa di marketing è stato chiaramente visibile nei risultati finanziari dell'azienda. Nel

[25]Si veda l'istruttivo saggio di: Benedikt Evans, "Abstracting Ai", in: Benedict's Newsletter: No. 528 (20. Feb. 2024).
[26]Per questo approccio "Designing for Dialogue", si veda : Your Organization Isn't Designed to Work with GenAI (hbr.org).
[27]Si veda : Giacomelli, G. (2024), Beyond "humans in the loop": reliable AI in enterprise workflows (linkedin.com).
[28]Rick et al (2023), Supermind Ideator: Exploring generative AI to support creative problem-solving (arxiv.org).
[29]Il GPT Store non è l'"app store" di ChatGPT, ma è comunque importante per i marketer (econsultancy.com).
[30]Il CEO di Coca-Cola: l'innovazione per il "vantaggio competitivo" (marketingweek.com).

complesso, questa fase di analisi dei bisogni e di riflessione concettuale e strategica sull'IA generativa è fondamentale per il successo dell'introduzione delle tecnologie linguistiche, in quanto costituisce la base per tutte le fasi successive e garantisce che l'introduzione della tecnologia sia adattata alle esigenze e agli obiettivi specifici dell'azienda. In quanto segue, partiamo dal presupposto che il management abbia già effettuato un'analisi delle esigenze e fissato gli obiettivi.

2.2 Dipendenze strategiche: pensare al potere di mercato

Una volta che un'azienda ha deciso di utilizzare l'IA generativa, deve scegliere tra l'utilizzo dell'IA generativa come prodotto cloud di un fornitore terzo (*"Artificial Intelligence as a Service*", AIaaS) e la creazione e l'installazione *in sede*: una scelta cruciale per le aziende, poiché ha un impatto diretto sulla scalabilità e sulla flessibilità delle soluzioni di IA[31]. L'AIaaS consente alle aziende di accedere ad algoritmi e risorse di calcolo avanzate (ad esempio tramite Microsoft Azure) senza dover mantenere una propria infrastruttura, come i server GPU dedicati. Al contrario, la scelta di strutture di IA interne può consentire un allineamento più stretto con le esigenze specifiche dell'azienda, che in genere si traduce in una maggiore efficienza e precisione delle applicazioni di IA. Oltre alla scelta strategica discussa in questa sezione, le PMI dovrebbero considerare anche i costi energetici e i requisiti di manutenzione (sezione 2.9), nonché la protezione dei dati (sezione 2.7) quando effettuano questa scelta.

Non disponendo di grandi *server* e chip specializzati per gestire i propri modelli linguistici su larga scala, come le grandi aziende tecnologiche globali (GAFAM) e le *start-up* che finanziano[32], le PMI devono scegliere tecnologie che possano essere facilmente adattate e che offrano la possibilità di aggiungere ulteriori funzionalità o capacità, se necessario. Inoltre, le tecnologie linguistiche scelte devono essere scalabili e flessibili, in modo da poter tenere il passo con la crescita e le mutevoli esigenze dell'azienda.

Questi due fattori depongono a favore dell'uso di LLM pre-addestrati. Questi possono essere proprietari e a pagamento (come gli ultimi modelli GPT di OpenAI) o gratuiti e *open-source* (come il modello Llama di Meta). È importante riconoscere che non tutti i modelli gratuiti sono automaticamente "veramente" *open-source* e possono non avere restrizioni d'uso; piuttosto, sono le condizioni in cui viene concesso l'accesso a essere cruciali: I modelli possono essere completamente chiusi (non disponibili per chiunque al di fuori dell'organizzazione che li ha sviluppati), resi disponibili tramite un'interfaccia web (ad esempio, l'API di GPT-4), offrire solo l'accesso basato su *cloud*, fornire l'accesso alla messa a punto (GPT-3 di OpenAI), rivelare i loro pesi (come Stability AI's Stable Diffusion e Meta's Llama 2), o essere disponibili con tutti i pesi, i codici e i dati[33]. Per le PMI, le considerazioni di cui sopra mostrano che le ultime categorie, spesso raggruppate sotto il termine "modelli di base aperti", sono particolarmente appropriate, in quanto i modelli sono pubblicati in modo trasparente e con pesi ampiamente disponibili.

Un altro criterio importante è la performance dei modelli. Poiché si tratta della chiave per un prodotto o servizio competitivo nei mercati a valle in cui opera la PMI, anche piccole differenze

[31]Bitkom (2024), L'AI generativa nell'impresa. Questioni legali relative all'uso dell'intelligenza artificiale generativa nelle imprese, pag. 16, 41.

[32]Von Thun (2024), Euractiv - L'UE non deve aspettare l'AI Act per agire - Open Markets Institute; Küsters e Kullas (2024), cep - European Policy Centre.

[33]Bommasani (2023), Considerations for Governing Open Foundation Models, Governing-Open-Foundation-Models.pdf (stanford.edu).

giocano un ruolo importante. A questo proposito, le varie metriche e indagini di settore sviluppate dai ricercatori delineano un quadro chiaro: il modello GPT-4 sviluppato da OpenAI è (attualmente) il leader del settore e batte tutti gli altri LLM, sia nei benchmark convenzionali che nei test progettati per essere valutati dagli esseri umani[34]. Recentemente, tuttavia, i modelli Claude 3 (Opus) e Gemini Ultra (Google) sono riusciti a recuperare terreno[35], lasciando alle aziende un certo margine di manovra. Ci sono poi modelli *open source* come Llama di Meta e, più recentemente, Mistral AI dalla Francia. Piuttosto che cercare di costruire da sole modelli costosi o aspettare che fornitori nazionali esclusivi lancino modelli equivalenti, le PMI europee dovrebbero quindi rivolgersi a questi modelli consolidati il prima possibile. Se un modello proprietario (come GPT-4) viene scelto rispetto a un modello *open source* a causa delle differenze di prestazioni, tuttavia, l'uso dei dati acquisiti dal fornitore di AI (come OpenAI) dovrebbe essere escluso da un accordo contrattuale o dalla scelta di una licenza specifica.

Ma questo non crea una dipendenza a lungo termine e, a causa dell'iniqua distribuzione del potere di mercato, dipendenze sfavorevoli ed effetti di preclusione? Per quanto riguarda l'uso dell'IA generativa nelle aziende, Bitkom mette in guardia contro "l'emergere di una dipendenza da fornitori di servizi esterni, ad esempio a causa della divulgazione di *know-how* o dati, nonché dei costi conseguenti (ad esempio a causa di aggiornamenti, manutenzione, cambio di fornitore di servizi esterni in un secondo momento)"[36]. Tuttavia, a differenza degli sviluppi precedenti nel mercato digitale, le dipendenze strategiche nell'area della tecnologia dei modelli linguistici dovrebbero essere molto meno significative, poiché i modelli sono relativamente facili da sostituire grazie all'accesso attraverso il linguaggio naturale, riducendo il potere di mercato dei principali sviluppatori. Come ha osservato di recente un esperto di IA: *"È notevole che si possa controllare un software dai parametri multimiliardari che si basa su centinaia di gigabyte di dati di input con un semplice inglese* "[37]. Questa accessibilità attraverso un linguaggio (relativamente) semplice significa che un piccolo numero di dipendenti con un background tecnico rudimentale può già interagire con sistemi software sofisticati e influenzarne il funzionamento. In passato, invece, era necessaria una conoscenza approfondita di linguaggi di programmazione come l'assembler o la manipolazione della memoria per ottenere un livello di controllo paragonabile.

Per le PMI, questo cambiamento nell'ecosistema dell'IA ha implicazioni significative per i potenziali costi di transizione in caso di aumento dei prezzi o di modelli obsoleti. La democratizzazione del controllo del software attraverso il linguaggio naturale riduce la necessità di competenze informatiche specialistiche e rende l'adozione di nuove tecnologie più facile ed economica per le PMI. Questo non solo aumenta la loro flessibilità nell'integrare soluzioni innovative, ma livella anche il campo di gioco con i concorrenti più grandi, il che può accelerare la trasformazione digitale e favorire un ambiente di mercato più competitivo. Ciò suggerisce che le PMI dovrebbero valutare i modelli di base aperti già stabiliti sulla base dei risultati dei test e delle risorse necessarie per implementarli, senza preoccuparsi troppo delle dipendenze successive. È importante selezionare il modello individuale giusto per assicurarsi che soddisfi le esigenze specifiche dell'azienda e possa

[34]Rapporto sullo Stato dell'IA 2023, Benvenuti al Rapporto sullo Stato dell'IA 2023. Si vedano anche le valutazioni attuali in: Chatbot Arena: Benchmarking LLMs in the Wild with Elo Ratings | LMSYS Org.

[35]Si veda l'analisi comparativa in : Warren (2024), Putting GPT-4's new rivals to the test (exponentialview.co).

[36]Bitkom (2024), L'AI generativa nell'impresa. Questioni legali relative all'uso dell'intelligenza artificiale generativa nell'impresa, pag. 16.

[37]Citato da : La brillante e complicata semplicità di ChatGPT (exponentialview.co).

contribuire efficacemente al raggiungimento degli obiettivi prefissati. Ma la rapidità della selezione è ancora più importante: quanto prima inizia l'introduzione dell'IA generativa, tanto più tempo e spazio ci sono per la necessaria sperimentazione.

2.3 *Fine-Tuning* e RAG: personalizzare l'IA con le proprie fonti di dati

I LLM pre-addestrati sono oggi ampiamente disponibili sul web (ad esempio tramite la popolare piattaforma HuggingFace) e sono considerati eccellenti tecnologie "generiche" che richiedono pochi o nessun esempio specifico per svolgere attività complesse, dalla stesura di comunicati stampa e relazioni aziendali più brevi alla creazione di grafici, presentazioni o persino applicazioni. Meta, Google e Mistral hanno creato famosi modelli pre-addestrati, facili da scaricare o da utilizzare tramite API. Ma per integrare con successo questi modelli linguistici pre-addestrati nei processi aziendali esistenti delle PMI, è necessario adattarli all'applicazione specifica. A questo scopo sono disponibili due tecniche in particolare: il *fine-tuning* e l'*Retrieval-Augmented Generation* (RAG). Entrambe consentono di ottimizzare i LLM per il contesto aggiungendo altre fonti di dati, ma ognuna presenta vantaggi e svantaggi propri[38].

La messa a punto di modelli linguistici di grandi dimensioni consiste nell'adattare LLM già addestrati su un set di dati generale, addestrandoli su un set di dati più piccolo e specifico, in modo che siano più adatti a un particolare dominio. Un buon esempio è "LEGAL-BERT", che ha ottimizzato il noto modello linguistico BERT per il dominio legale e l'applicazione della tecnologia giuridica, addestrandolo ulteriormente su diversi testi giuridici (ad esempio, leggi, atti giudiziari, contratti)[39]. Questa tecnica consente quindi ai modelli di adattare le loro ampie conoscenze generali ai requisiti differenziati di determinate applicazioni, migliorando le loro prestazioni per compiti specifici e la loro accuratezza. Per le PMI, la messa a punto dell'LLM offre l'opportunità di adattare una tecnologia linguistica IA all'avanguardia alle proprie esigenze aziendali, senza dover sostenere i costi elevati dello sviluppo e della formazione di modelli di base completamente nuovi. Grazie alla messa a punto degli LLM su dati interni e set di documenti che riflettono il loro specifico contesto aziendale, le PMI possono ottenere risultati più accurati ed efficaci in aree quali l'automazione del servizio clienti, il marketing personalizzato e la creazione di contenuti, consentendo loro di distinguersi nel settore (poiché nessun altro ha accesso ai loro dati interni).

La *Retrieval-Augmented Generation* (RAG) migliora l'efficacia dei LLM integrando fonti esterne di informazioni dopo l'apprendimento (nella "fase di recupero"). In particolare, l'algoritmo cerca attivamente frammenti di informazioni rilevanti in risposta alle domande dell'utente e li recupera in modo tale che possano essere sintetizzati da modelli linguistici generativi, come quelli basati su trasformatori come GPT, per produrre risposte coerenti e contestualmente rilevanti[40]. Si prevede che l'accesso ai fatti più recenti e rilevanti migliorerà in modo significativo le risposte, ad esempio se utilizzato nei chatbot Q&A per i clienti. La RAG promuove anche la trasparenza e l'affidabilità, consentendo agli utenti di verificare le fonti delle informazioni utilizzate dall'IA. Nonostante il suo potenziale, il concetto di RAG è ancora sconosciuto in molte aziende, poiché se ne parla solo da un tempo relativamente breve[41]. Se da un lato questo approccio migliora l'adattabilità e l'accuratezza

[38]Per un confronto, vedere :Generazione aumentata di recupero: mantenere i LLM rilevanti e attuali - Stack Overflow.

[39]Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras e Ion Androutsopoulos. 2020. LEGAL-BERT: Muppet direttamente dalla scuola di legge. In Findings of the Association for Computational Linguistics: EMNLP 2020, pagine 2898-2904, online. Associazione per la linguistica computazionale.

[40]Si veda: 12 strumenti/software di Retrieval Augmented Generation (RAG) in '23 (aimultiple.com).

[41]Si veda ad esempio Andrew Ng: My Daily Note Taking Device: reMarkable 2 (2023) (youtube.com).

dei LLM, superando i limiti di conoscenza statica insiti in questi modelli, dall'altro comporta maggiori requisiti computazionali, latenze più lunghe e richieste più complesse. Per questo motivo, è raccomandato (per ora) solo per i casi di applicazione in cui la velocità di inferenza e il consumo di risorse non sono troppo importanti. Sebbene la costruzione di un modello RAG sia relativamente semplice, secondo una recente revisione sono necessari adattamenti significativi e una comprensione relativamente profonda del dominio applicativo per ottenere un'applicazione robusta e affidabile[42]. Infine, va sottolineato che i modelli RAG non sono una panacea e soffrono ancora di problemi metodologici[43]. Una recente valutazione dei modelli RAG in diverse aree cliniche ha dimostrato che l'inclusione dei RAG ha ridotto significativamente il numero di errori, ma che anche nel modello migliore (GPT-4 RAG), fino al 30% delle asserzioni non era supportato da una delle fonti indicate[44].

Insieme, Fine-Tuning e RAG consentono di sviluppare applicazioni LLM su misura per le PMI, specializzate in aree specifiche e che utilizzano la conoscenza contestuale. In futuro, questi assistenti di intelligenza artificiale aziendale svolgeranno un ruolo sempre più importante nel rendere più efficienti i processi di lavoro, sfruttando meglio le conoscenze interne. Un buon esempio è GitHub Copilot, che utilizza l'ambiente di programmazione esistente come base di conoscenza per contestualizzare e rispondere meglio alle richieste dei programmatori interni. Possiamo aspettarci che d'ora in poi "copiloti" simili vengano formati da molte aziende.

2.4 Sviluppare le competenze interne di PNL: (Prompt) Design

Le PMI devono sviluppare una conoscenza di base dei modelli linguistici disponibili e dell'elaborazione del linguaggio naturale (NLP *Natural Language Processing*). Ciò comporta non solo una valutazione dei diversi modelli in termini di capacità, limiti e costi potenziali di follow-up (vedi sopra), ma soprattutto lo sviluppo di competenze interne di *prompting*. I prompt nei modelli di IA generativa sono input testuali che controllano l'output del modello e vanno da semplici domande a compiti dettagliati[45]. Nei modelli che generano immagini, come DALL-E, i prompt sono spesso descrittivi, mentre nei modelli linguistici, come GPT-3, possono variare da semplici domande a problemi complessi. Grazie all'attuale tendenza verso i cosiddetti LLM multimodali, le istruzioni testuali possono ora essere generalmente integrate con immagini, testo aggiuntivo o audio, di cui il sistema di intelligenza artificiale tiene conto quando crea i formati corrispondenti. Poiché lo sviluppo di suggerimenti appropriati richiede un notevole investimento in termini di tempo e personale, questi sono probabilmente un segreto commerciale degno di essere protetto nel senso legale del termine[46].

Forse la strategia di *prompting* più nota è quella di scomporre i compiti complessi nelle loro parti componenti. Finora, l'evidenza scientifica sostiene questa strategia "passo dopo passo": sembra che

[42]Fatehkia et al (2024), [2402.07483] T-RAG: Lessons from the LLM Trenches (arxiv.org).
[43]A questo proposito, si veda la posizione scettica di Gary Marcus: Non, RAG probabilmente non salverà la situazione attuale (substack.com).
[44]Wu et al (2024), [2402.02008] How well do LLMs cite relevant medical references?Un quadro di valutazione e analisi (arxiv.org).
[45]Per una panoramica, si veda: [2401.14423] Prompt Design and Engineering: Introduction and Advanced Methods (arxiv.org).
[46]Bitkom (2024), L'AI generativa nell'impresa. Questioni legali relative all'uso dell'intelligenza artificiale generativa nelle imprese, pag. 41.

la qualità del lavoro migliori quando al modello viene chiesto di scomporre un compito nelle sue parti componenti. Uno studio empirico è stato in grado di dimostrare che questo *primer* della "catena del pensiero" (CoT) può guidare con successo i modelli linguistici attraverso processi di pensiero a più livelli per ottenere prestazioni elevate in compiti complessi come l'aritmetica e il ragionamento simbolico.[47] I modelli linguistici sono quindi in grado di ottenere buone prestazioni in compiti senza costosi esempi codificati, aggiungendo un invito a pensare ai problemi passo dopo passo (nell'originale inglese, ad esempio: *"Let's think step by step* "). Studi successivi hanno dimostrato che i LLM possono migliorare significativamente la produttività e la qualità del processo di generazione delle idee quando viene utilizzato questo tipo di CoT-Prompting.[48] Le PMI possono utilizzare tali tecniche di *prompting* per sbloccare meglio le capacità cognitive dei LLM per i loro scopi specifici.

Alcune delle tecniche di stimolo più efficaci sembrano talvolta controintuitive e sono il risultato di esperimenti sorprendenti, il che dimostra l'importanza di un esperto del settore che abbia familiarità con gli sviluppi attuali. Alcuni studi hanno dimostrato, ad esempio, che gli appelli alle emozioni nel *prompt* (ad esempio "Questo è personalmente molto importante per me") portano a risultati significativamente migliori. Nell'articolo in questione, i ricercatori hanno testato modelli linguistici di *prompt* con e senza emozioni aggiuntive e hanno scoperto che questi ultimi hanno portato a un miglioramento medio del 10,9% nelle aree della performance, della veridicità e del senso di responsabilità.[49] Non si comprende ancora il meccanismo sottostante, ma funziona. Lo studio ha testato modelli importanti come ChatGPT, Llama 2 e altri LLM, quindi è probabile che questo valga anche per i modelli linguistici interni delle PMI, basati su questi modelli pre-addestrati.

Oltre alla posizione eccezionale dei *prompt designer* nell'integrazione delle LLM, un secondo gruppo di professionisti, più inaspettato, svolge un ruolo speciale nella pratica imprenditoriale: i classici online-designer. Nell'era dell'IA generativa, la competitività e l'attrattiva per i clienti dipenderanno non solo dalle capacità tecnologiche, ma anche dalla qualità e dalle capacità di progettazione delle aziende[50]. L'ecosistema digitale emergente, in cui i modelli linguistici stanno diventando sempre più la nuova piattaforma e infrastruttura di accesso a Internet[51], premia chi offre interfacce utente di qualità superiore e un'integrazione senza soluzione di continuità, caratteristiche che sono il segno distintivo di designer qualificati. Per rimanere competitive, le PMI devono quindi dare priorità al reclutamento e alla promozione dei talenti del design. Le *start-up*, in particolare, possono ottenere un vantaggio competitivo concentrandosi sul design innovativo per creare esperienze utente nuove e trasformative, differenziandosi così da un mercato consolidato in cui i concorrenti sono tecnologicamente abili ma scarsi nel design.

[47]Kojima et al (2022), [2205.11916] Large Language Models are Zero-Shot Reasoners (arxiv.org).
[48]Meincke et al. (2024), Prompting Diverse Ideas: Increasing AI Idea Variance, SSRN.
[49]Li et al (2023), Large Language Models Understand and Can Be Enhanced by Emotional Stimuli, 2307.11760.pdf (arxiv.org).
[50]Questo è l'argomento di : Belsky (2024), The Era of Abstraction & New Creative Tensions (implications.com).
[51]Sulla futura influenza dei modelli linguistici in generale, si veda: [2305.07961] Leveraging Large Language Models in Conversational Recommender Systems (arxiv.org). Sull'impatto di un mondo digitale sempre più astratto, si veda :The Era of Abstraction & New Creative Tensions (implications.com).

2.5 Il caso delle allucinazioni: adattare il tasso di errore dell'IA alla propria tolleranza all'errore

L'integrazione dei LLM nei processi aziendali delle PMI non è priva di sfide, in particolare per quanto riguarda la tendenza di questi modelli alle cosiddette allucinazioni, ovvero la generazione di informazioni plausibili, ma in realtà false o assurde.[52] Da dove derivano questi errori? I grandi modelli linguistici non sono stati progettati per il recupero di informazioni esterne e le loro prestazioni sono ulteriormente limitate dal volume e dalla tempestività dei dati con cui sono stati addestrati. Quando i LLM non dispongono di informazioni sufficienti per fornire una risposta fondata, costruiscono risposte basate su input precedenti. Questo rappresenta un rischio significativo per le PMI, poiché le imprecisioni nella comunicazione con i clienti, nella generazione dei contenuti o nell'analisi dei dati possono portare alla diffusione di informazioni fuorvianti, minando la fiducia dei clienti e danneggiando potenzialmente la reputazione del marchio. Da un punto di vista legale, occorre tenere presente che tali errori di IA incorporati in prodotti o servizi potrebbero portare a violazioni contrattuali, responsabilità e multe.[53]

Oltre alla possibilità di errori "involontari", l'integrazione di *bot* vocali apre anche un significativo potenziale di abuso, ad esempio da parte di aggressori esterni. Poiché la funzionalità LLM può essere modulata in modo flessibile tramite messaggi in linguaggio naturale (piuttosto che tramite codice), è vulnerabile a messaggi mirati che consentono agli aggressori di aggirare le istruzioni e i controlli iniziali. I ricercatori hanno descritto vettori di attacco che consentono di abusare delle applicazioni embedded LLM da remoto, inserendo *prompt* mirati in dati a cui è probabile accedere*(attacchi indiretti di prompt injection*).[54] Un gruppo di hacker vicini alla Russia ha rivendicato la responsabilità degli attacchi che hanno temporaneamente disabilitato ChatGPT alla fine del 2023.[55] Ciò ha comportato interruzioni parziali e, secondo quanto riferito, tassi di errore più elevati tra gli utenti di ChatGPT. Se una PMI fa affidamento sull'accesso continuo ai modelli GPT di OpenAI, tali attacchi potrebbero interrompere i processi aziendali interni. Infine, i ricercatori sono riusciti a dimostrare che è possibile costruire "backdoor" nei LLM, cioè addestrarli ad adottare un comportamento ingannevole e, ad esempio, a eseguire codice maligno in un secondo momento.[56] Questo comportamento ingannevole, definito dai ricercatori "agenti dormienti", persiste anche dopo le procedure di formazione standard sulla sicurezza. Questi risultati ed esperienze di ricerca evidenziano la necessità di contromisure efficaci per proteggere i sistemi gestiti da LLM, suggerendo al contempo che le misure esistenti non sono ancora sufficienti.

Tuttavia, negli ultimi due anni, programmatori e utenti hanno scoperto molte misure che possono contribuire a mitigare il livello di “allucinazioni” e altri errori, nonché il potenziale di abusi esterni.[57] Ad esempio, il semplice fatto di fornire esempi concreti (testo, codice o dati) nella richiesta a un LLM può già migliorare significativamente la rilevanza e la qualità dell'output. Inoltre, la tecnica RAG citata

[52] Esplorazione di grandi modelli linguistici (LLM): AI e allucinazioni | ZS.
[53] Bitkom (2024), L'IA générative dans l'entreprise. Questioni giuridiche relative all'uso dell'intelligenza artificiale generativa nell'impresa, pag. 44.
[54] [2302.12173] Non è quello per cui hai firmato:Compromissione di applicazioni integrate in LLM del mondo reale con l'iniezione indiretta di prompt (arxiv.org).
[55] Hacker legati alla Russia rivendicano il merito della fuga di notizie su OpenAI di questa settimana - BNN Bloomberg - OECD.AI.
[56] [2401.05566] Agenti dormienti:Formazione di LLM ingannevoli che persistono attraverso l'addestramento alla sicurezza (arxiv.org).
[57] Per una buona panoramica, si veda : Come ridurre l'allucinazione in un Large Language Model (LLM)?(linkedin.com).

in precedenza riduce il rischio di risultati imprecisi o errati, poiché cerca informazioni rilevanti in fonti credibili e le utilizza per integrare le risposte del LLM.[58] Nella letteratura sul *prompted design* (si veda la sezione 2.4), il rischio di errore nei modelli linguistici è ulteriormente migliorato dall'integrazione di alcune componenti tecniche ("strumenti", "connettori" o "competenze").[59] Queste estensioni ai normali LLM consentono allo strumento linguistico di accedere e interagire con fonti di dati esterne e di svolgere compiti che vanno oltre le sue capacità integrate. Questo potenziale di utilizzo ampliato rende la tecnologia dei modelli linguistici più interessante per un'ampia gamma di PMI europee, in quanto i compiti che consente di svolgere vanno dal semplice reperimento di dati a complesse interazioni con banche dati o API, sintesi di testi o traduzione vocale sensibile al contesto. In futuro, i LLM potrebbero persino imparare da soli quando e come chiamare e utilizzare strumenti esterni tramite semplici API.[60] In questo modo, diventeranno agenti di intelligenza artificiale che potranno anche essere integrati in processi e prodotti fisici (si veda anche la sezione seguente).

In sintesi, non esiste ancora un solido quadro di verifica e validazione che le PMI possano implementare rapidamente per eliminare completamente gli effetti negativi delle allucinazioni LLM e di altre categorie di errori. Ciò può avere conseguenze legali: un tribunale canadese ha recentemente stabilito che Air Canada ha dovuto pagare i danni a un passeggero perché il *chatbot* del servizio clienti alimentato dall'intelligenza artificiale gli aveva dato consigli fuorvianti e il passeggero ha dovuto pagare quasi il doppio del biglietto aereo come risultato.[61] In una certa misura, tali errori rimarranno sempre possibili nonostante i progressi dell'IA generativa, poiché i modelli operano in ultima analisi in modo probabilistico, il che significa che si limitano a prevedere una determinata sequenza di *token* senza disporre di un modello sottostante di contesto (il che spiega perché l'output può talvolta essere diverso nonostante un *prompt* identico). È quindi essenziale che le PMI adattino il tasso di errore dell'IA (e il potenziale di abuso) alla propria tolleranza interna agli errori, in modo che, in caso di allucinazione, l'integrità e l'affidabilità dell'azienda non siano compromesse. In altre parole, prima di qualsiasi utilizzo, è importante valutare se un'allucinazione in questa specifica area avrebbe conseguenze significative o potrebbe essere facilmente corretta. Inoltre, è possibile tutelarsi attraverso disposizioni contrattuali relative a garanzia, responsabilità ed esonero.[62]

2.6 Agenti incorporati: integrazione in processi, prodotti e servizi

Nella letteratura sull'IA, il termine "agente" descrive un sistema in grado di svolgere determinati compiti in modo autonomo.[63] Uno degli aspetti più notevoli dei modelli linguistici di IA è la loro capacità di utilizzare strumenti software esterni per raggiungere obiettivi predefiniti. Proprio come gli esseri umani scrivono codice o utilizzano software al di là delle loro capacità immediate, i LLM possono imitare questo processo per svolgere determinati compiti. Ad esempio, possono essere addestrati a riconoscere quando è utile utilizzare un'interfaccia di programmazione, elaborare i dati ricevuti e adattare le loro azioni di conseguenza.[64] Ciò rende possibile lo sviluppo di agenti di

[58] 12 Strumenti / Software di Retrieval Augmented Generation (RAG) en '23 (aimultiple.com).

[59] Vedi: [2401.14423] Prompt Design and Engineering : Introduction and Advanced Methods (arxiv.org).

[60] Vedi: [2302.04761] Toolformer: Language Models Can Teach Themselves to Use Tools (arxiv.org).

[61] Air Canada ordina il pagamento delle tasse ai passeggeri dopo che il chatbot parla di sconti sui trasporti (gizmodo.com).

[62] Bitkom (2024), Generative AI in the enterprise. Questioni legali relative all'uso dell'intelligenza artificiale generativa nelle imprese, pag. 49.

[63] Per una panoramica, si veda: [2401.14423] Prompt Design and Engineering: Introduction and Advanced Methods (arxiv.org).

[64] [2302.04761] Toolformer: I modelli linguistici possono insegnare a usare gli strumenti (arxiv.org).

intelligenza artificiale avanzati che utilizzano software diversi per migliorare le loro capacità o colmare le lacune. Questi agenti di intelligenza artificiale sono progettati per interagire sia con gli utenti che con l'ambiente circostante e per prendere decisioni informate in base agli input ricevuti e agli obiettivi predefiniti.[65] Sono destinati a compiti che richiedono un certo grado di autonomia nel processo decisionale e nella risoluzione dei problemi, al di là della semplice generazione di risposte.

Sebbene l'implementazione di questo modello non sia ancora pronta per l'uso pratico, gli esperti prevedono che ulteriori sviluppi di agenti basati su LLM diventeranno sempre più rilevanti dal punto di vista commerciale.[66] OpenAI, ad esempio, sta lavorando a un agente di intelligenza artificiale che prende il controllo del dispositivo dell'utente e consente al software di eseguire clic, battute e altre azioni.[67] Analogamente, Apple sta lavorando per portare l'IA generativa sui dispositivi mobili.[68] Google ha già integrato i suoi modelli linguistici Gemini in molti dei suoi servizi, tra cui Android, l'applicazione Google per iOS e Gmail.[69] In questo caso, l'introduzione di "personalità" basate su LLM ha contribuito allo sviluppo di molti *bot* che interagiscono autonomamente con gli utenti e possono simulare amicizie o interessi meglio che mai.[70] In futuro, questi assistenti più potenti potrebbero essere combinati con l'intelligenza artificiale *text-to-speech*. Da un punto di vista legale, è interessante notare che un agente di IA generativa non ha una capacità legale o commerciale propria, ma può contribuire come aiuto alle prestazioni in alcuni processi, come l'automazione di compiti di routine e l'analisi dei dati (la responsabilità rimane all'operatore).[71]

Cosa significa questo per le aziende europee che fabbricano prodotti fisici? Le PMI non dovrebbero considerare le nuove tecnologie linguistiche come tecnologie puramente testuali che rimangono sugli schermi dei computer, ma dovrebbero pensare per tempo a come integrarle sempre più nei prodotti e nei servizi fisici del loro settore. Ad esempio, potremmo immaginare un agente AI basato su LLM che abbia accesso a un'API per lo *shopping*, ottenga informazioni da fonti esterne (ad esempio un comparatore di prezzi) e poi effettui autonomamente determinati acquisti tramite l'API sulla base di queste informazioni (ad esempio, si faccia consegnare il prodotto più economico disponibile). Allo stesso modo, le PMI potrebbero utilizzare agenti di intelligenza artificiale per ottimizzare i loro sistemi di gestione della catena di approvvigionamento o offrire assistenti che guidino i clienti nel processo di acquisto e consentano configurazioni personalizzate dei prodotti. Questo tipo di interazione uomo-macchina può non solo aumentare la fedeltà dei clienti, ma anche fornire informazioni sulle loro preferenze, che a loro volta possono essere utilizzate come dati di formazione per lo sviluppo futuro dei prodotti e il miglioramento dei modelli.

Tuttavia, un eccessivo affidamento agli agenti basati su LLM per i processi decisionali critici, senza un adeguato monitoraggio, può portare a errori strategici, motivo per cui dovrebbero essere testati attentamente e non utilizzati per funzioni critiche. Un esempio drammatico evidenzia questo rischio:

---

[65] [2401.14423] Prompt Design and Engineering: Introduction and Advanced Methods (arxiv.org).

[66] Questa valutazione si basa principalmente su interviste con esperti durante l'8° Open European Dialogue di Helsinki, vedi: openeuropeandialogue.org/download-file/2296/. Si veda anche la discussione ottimistica in : Lazar (2024), Can philosophy help us get a grip on the consequences of AI? | Aeon Essays.

[67] OpenAI trasforma il campo di battaglia dell'IA in un software che gestisce i dispositivi e automatizza i compiti - The Information.

[68] Apple potenzia i piani per portare l'IA generativa sugli iPhone (ft.com).

[69] Gemini di Google è ora in tutto.Ecco come provarlo (MIT Technology Review).

[70] [2303.06135] Rewarding Chatbots for Real-World Engagement with Millions of Users (arxiv.org). Vedi anche: My AI lover | Psyche Films.

[71] Bitkom (2024), L'AI generativa nell'impresa. Questioni legali relative all'uso dell'intelligenza artificiale generativa nelle imprese, pag. 59.

gli scienziati hanno studiato l'uso di agenti basati su LLM in giochi strategici di tipo militare e hanno trovato "forme e modelli di *escalation* difficili da prevedere".[72] Hanno scoperto che i modelli sviluppavano dinamiche negative affidandosi a "giustificazioni minacciose" come le tattiche di primo attacco. In generale, la ricerca ha dimostrato che è possibile modificare l'orientamento morale ed etico di un modello di IA, anche per i modelli più potenti come il GPT-4.[73] Alla luce di tali rischi, gli agenti autonomi nei modelli linguistici non dovrebbero quindi essere utilizzati in prima istanza per prendere importanti decisioni strategiche.

2.7 Condizioni legali: protezione dei dati e delle conoscenze, sfruttamento della legge sull'IA

Nonostante l'euforia, le PMI devono tenere conto di alcuni quadri giuridici quando integrano i LLM nei loro processi aziendali. Sia l'uso dei dati immessi nei modelli di IA che l'uso dei risultati dell'IA riguardano una serie di questioni legali, tra cui il diritto d'autore, la protezione dei dati, le questioni di responsabilità e il diritto del lavoro.[74] La protezione della privacy, in particolare, è oggetto di un vivace dibattito volto ad aggiornare le norme esistenti per tenere il passo con il panorama giuridico di un mondo sempre più incentrato sui dati.[75] Le PMI con sede in Europa devono tenere conto del Regolamento generale sulla protezione dei dati (GDPR) dell'UE, la legge sull'IA dell'UE i cui negoziati finali si sono conclusi alla fine di dicembre 2023, nonché di altre normative nazionali e internazionali (in Germania, ad esempio, la legge federale sulla protezione dei dati e la legge sulla protezione dei segreti commerciali). Queste dovrebbero garantire la protezione dei dati personali e dei segreti commerciali e che l'uso dell'IA generativa, ad esempio sotto forma di *chatbo*t, avvenga in modo trasparente. Di seguito vengono analizzate in dettaglio tre aree problematiche: In primo luogo, i dati sensibili possono fuoriuscire dai sistemi e, se necessario, diventare visibili agli aggressori o anche a parti non coinvolte ("data leakage"). In secondo luogo, esistono ancora problemi di *copyright* relativi alle fonti da cui proviene la tecnologia dei modelli linguistici. In terzo luogo, nuovi obblighi - ma anche interessanti diritti - derivano dalla legislazione europea sull'intelligenza artificiale.

Gli LLM possono essere utilizzati per spiare i dati privati di persone o aziende.[76] Utilizzando metodi relativamente semplici, come chiedere agli utenti di ripetere all'infinito una parola come "poesia", i ricercatori sono riusciti a far sì che ChatGPT rivelasse involontariamente gran parte dei suoi dati di addestramento.[77] Per le PMI, la questione è quindi come scoprire in modo sicuro nuovi casi d'uso di LLM basati su dati interni; anche perché tutto ciò che viene caricato su servizi commerciali di LLM potrebbe potenzialmente essere registrato come futuri dati di formazione.[78] Nel frattempo OpenAI ha risposto ad alcune vulnerabilità note e ha adottato misure per impedire agli aggressori di inviare inconsapevolmente i dati degli utenti a server esterni.[79] Nonostante questi miglioramenti, le

[72] [2401.03408] Escalation Risks from Language Models in Military and Diplomatic Decision-Making (arxiv.org). Vedi anche: Il GPT-5 potrebbe rivoluzionare la strategia militare? (substack.com).

[73] [2311.05553] Rimozione delle protezioni RLHF in GPT-4 via Fine-Tuning (arxiv.org).

[74] Bitkom (2024), IA generativa nell'impresa. Questioni legali relative all'uso dell'intelligenza artificiale generativa nell'impresa.

[75] Le leggi esistenti e proposte in materia di privacy regolano implicitamente lo sviluppo dell'IA, ma sono considerate dagli esperti insufficienti a regolamentare in modo sufficiente l'attuale corsa ai dati. Per una panoramica, si veda: King e Meinhardt (2024), White-Paper-Rethinking-Privacy-AI-Era.pdf (stanford.edu).

[76] Tre modi in cui i chatbot AI sono un disastro per la sicurezza | MIT Technology Review.

[77] ChatGPT può far trapelare dati di addestramento e violare la privacy, dice DeepMind di Google | ZDNET.

[78] Usare l'open source per esperimenti di IA generativa più sicuri (mit.edu).

[79] Si veda l'analisi di :OpenAI inizia ad affrontare la vulnerabilità della fuga di dati di ChatGPT - Embrace The Red

preoccupazioni rimangono, in quanto le fughe di dati sono ancora possibili con alcuni metodi di attacco. Resta da vedere quanto queste misure saranno efficaci a lungo termine e se la sicurezza dei dati delle PMI potrà essere garantita.

Il rispetto del diritto d'autore è oggetto di un vivace dibattito accademico e di numerose battaglie legali ancora oggi irrisolte. La più nota è senza dubbio la denuncia presentata dal New York Times (NYT) contro Microsoft e OpenAI, che sostiene che i servizi di intelligenza artificiale come ChatGPT utilizzano illegalmente i contenuti del giornale. I querelanti chiedono che tutte le IA addestrate sui loro articoli vengano ritirate. Al centro della denuncia c'è l'accusa del Times secondo cui gli LLM sono "macchine copiatrici di massa" che producono "copie quasi esatte" di porzioni significative di articoli di NTY su richiesta.[80] Secondo gli osservatori del processo, sembra che la denuncia travisi il funzionamento degli LLM e utilizzi esempi selettivi che forniscono una narrazione moralmente accattivante, ma nessuna solida argomentazione legale.[81] In realtà, l'IA generativa non si basa su un algoritmo predefinito, ma su metodi statistici. In parole povere, possiamo dire che i modelli linguistici non imparano a memoria i testi originali, ma solo le probabilità. Il fatto che il modello a volte riproduca alcuni articoli del NYT quasi parola per parola è dovuto più che altro al fatto che questi testi sono stati spesso copiati o condivisi su Internet (e quindi fanno parte del modello statistico), oppure che sono molto specifici (dal punto di vista tematico e linguistico) e quindi possono essere attivati da "prompt" altrettanto specifici. Secondo gli informatici, quindi, non ha senso confrontare la misura in cui l'output di ChatGPT corrisponde esattamente agli articoli originali. Se la decisione dei giudici si concentra su questo punto, potrebbe, secondo questi osservatori, rendere più difficile la soluzione del problema di fondo, ossia la mancanza di partecipazione finanziaria degli autori alla conoscenza che hanno creato.[82] Analogamente, l'opinione prevalente in ambito giuridico non considera attualmente l'estrazione di informazioni da opere protette e l'adattamento dei valori di ponderazione della rete neurale su cui si basa una tecnologia di modelli linguistici di intelligenza artificiale come ChatGPT come una riproduzione punibile delle opere addestrate.[83] Anche se il tribunale si pronuncerà contro il NYT, le PMI dovrebbero tenere d'occhio questa questione legale, poiché avrà un impatto sui modelli che possono utilizzare e sulla possibilità di addestrare autonomamente i sistemi di IA utilizzando dati disponibili pubblicamente su Internet. Il Garante italiano per la protezione dei dati personali ha recentemente completato l'indagine su ChatGPT, che ha portato al divieto temporaneo del chatbot l'anno scorso, e ha riscontrato diverse violazioni delle norme sulla protezione dei dati.[84] Indagini sulle pratiche di OpenAI in materia di protezione dei dati sono in corso anche in Spagna, Francia e Germania. Lo sviluppo di politiche interne che rispettino le leggi vigenti in materia di protezione dei dati è quindi fondamentale per le PMI per mantenere la fiducia dei clienti a lungo termine nonostante l'uso dell'IA generativa.

Anche le nuove norme dell'UE in materia di IA, concordate dai negoziatori degli Stati membri alla fine del 2023 dopo intense discussioni, potrebbero contribuire a rafforzare la fiducia dei clienti.[85] Queste

[80] NYT_Complaint_Dec2023.pdf (nytimes.com)

[81] La causa del New York Times sul copyright contro OpenAI minaccia il futuro dell'IA e del fair use - Center for Data Innovation.

[82] La partita finale del copyright dell'IA generativa non sarà risolta dai tribunali (aisnakeoil.com).

[83] Bitkom (2024), L'AI generativa nell'impresa. Questioni legali relative all'uso dell'intelligenza artificiale generativa nell'impresa, pag. 38.

[84] ChatGPT : Garante privacy, notificato a OpenAI l'atto di contestazione... - Garante della privacy.

[85] Per una valutazione dell'accordo, si veda: cep - Centrum für europäische Politik: EU AI Act: A Milestone Met, But Key Challenges Remain in Standardisation and Competition. Il testo finale è disponibile qui: AM_Ple_LegConsolidated (europa.eu).

norme mirano a garantire che i modelli di IA siano utilizzati in Europa in modo etico, sicuro e rispettoso e che tutelino i diritti fondamentali. Il rispetto delle regole è obbligatorio per tutti i fornitori, distributori o operatori di sistemi e modelli di IA immessi sul mercato dell'UE.[86] I requisiti variano a seconda del livello di rischio e comprendono quattro categorie di rischio, da inaccettabile a minimo, ciascuna con obblighi specifici e scadenze che vanno da sei a 36 mesi. Ad esempio, mentre i filtri antispam rappresentano un rischio basso, i test di credito sarebbero considerati un rischio elevato in quanto comportano un rischio di discriminazione. All'IA generativa si applicano obblighi specifici, a seconda che si tratti o meno di un modello open source. Le PMI che intendono integrare la tecnologia dei modelli linguistici dovrebbero comprendere la categoria di rischio del loro sistema di IA e prepararsi fin da ora a rispettare le norme UE in materia.

In relazione all'argomento qui trattato, è particolarmente importante che tutti i modelli di base di IA (indipendentemente dal rischio) debbano soddisfare i requisiti di trasparenza prima di essere immessi sul mercato, ad esempio per quanto riguarda l'uso, l'architettura, i dati di formazione e altra documentazione tecnica.[87] Sono state introdotte norme più severe per i modelli di base di importanza sistemica (definiti provvisoriamente in termini di numero di FLOPS). Si tratta di modelli di base che vengono addestrati con grandi quantità di dati e la cui complessità e performance sono ben al di sopra della media, il che può portare alla propagazione di rischi sistemici lungo la catena del valore. Questi modelli di base di importanza sistemica (noti come sistemi di IA ad alto rischio) devono prima essere sottoposti a una procedura di valutazione della conformità (art. 8 e seguenti e art. 43 del Regolamento europeo sull'IA). Per le PMI che stanno pensando di integrare un chatbot basato sull'IA, è importante sapere che il Regolamento introduce nuove possibilità per la divulgazione e la tracciabilità dei contenuti generati artificialmente, nonché per informare gli utenti finali che hanno a che fare con un chatbot basato sull'IA (art. 50 del Regolamento europeo sull'IA). Poiché è probabile che le PMI facciano frequentemente uso di modelli di base forniti da sviluppatori esterni, spesso statunitensi (si veda la sezione 2.2), è importante che la normativa UE sull'IA formuli regole che consentano agli utenti successivi dei modelli di base di comprenderli meglio e di ottenere tutte le informazioni necessarie per un'implementazione sicura (art. 13 e 53 e seguenti del Regolamento UE sull'IA). Tuttavia, in caso di modifiche sostanziali del modello iniziale, può accadere che il fornitore che ha inizialmente immesso il sistema di IA sul mercato non sia più considerato tale - la responsabilità viene quindi trasferita all'utente del modello che ha apportato le modifiche.

Da una prospettiva globale che va oltre le singole normative, la legge sull'IA crea un sistema di *governance* eccessivamente complesso con un alto grado di incertezza giuridica. Come ha recentemente sottolineato Kai Zenner, che ha partecipato alle negoziazioni della legge sull'IA, questo mix di complessità e insicurezza giuridica, con molti termini legali indeterminati, "potrebbe aumentare significativamente i costi di conformità per i fornitori e gli utenti di IA". In particolare, le PMI e le *start-up* dell'UE potrebbero finire per ritenere troppo rischioso sviluppare o utilizzare l'IA o potrebbero essere costrette a ricorrere a costosi *audit* e certificazioni di terzi per evitare pesanti multe".[88] Per evitare questo scenario, le PMI dovrebbero seguire da vicino l'attuazione concreta della legge sull'IA con la serie di regolamenti di attuazione e atti delegati che saranno adottati in linea con l'entrata in vigore graduale della legge sull'IA tra il 2025 e il 2027. Inoltre, la legge sull'IA consente di

---

[86] Per una panoramica della legge sull'IA che aiuti le aziende a conformarsi al regolamento, vedere : Compliance AI Act - Feb 24 (wavestone.com). La sintesi qui presentata si basa su tale guida.

[87] Legge sull'IA: Consiglio e Parlamento cercano un accordo sulle prime norme mondiali sull'IA - Consilium (europa.eu)

[88] Alcune riflessioni personali sulla legge europea sull'intelligenza artificiale: una fine dolceamara (linkedin.com).

richiedere *sandbox* regolamentari (articoli 57 e seguenti del regolamento UE sull'IA). In questo modo, le PMI possono impegnarsi in uno stretto dialogo con le autorità nazionali competenti e testare e migliorare i loro sistemi di IA in condizioni reali e senza incertezze giuridiche. In ogni caso, quando integrano l'IA generativa, le PMI dovrebbero coinvolgere fin dall'inizio esperti legali e di conformità competenti, integrare la propria governance dell'IA nel processo di appalto e, se del caso, istituire un proprio sistema di gestione del rischio ai sensi della normativa europea sull'IA.[89]

2.8 Test interni: valutare la propria *AI-ness*

Prima che un sistema di IA venga pienamente implementato nell'attività quotidiana di un'azienda, il modello linguistico scelto deve essere accuratamente testato per garantirne l'accuratezza, l'affidabilità e l'efficacia. Ciò comporta test in condizioni reali e la valutazione dei modelli in base a specifici indicatori di performance. I controlli di qualità devono essere eseguiti regolarmente anche dopo l'implementazione, per garantire un livello elevato di prestazioni. In sostanza, l'obiettivo è quello di comprendere meglio le proprietà o il "carattere" del modello di IA utilizzato: un compito complesso, reso più difficile dal fatto che queste proprietà possono cambiare nel tempo o avere effetti collaterali indesiderati. Ad esempio, lo stile di comunicazione delle personalità guidate dall'IA, ad esempio nei colloqui con i clienti, è talvolta percepito dagli esseri umani come così genuino, professionale e premuroso che possono verificarsi effetti collaterali psicologici.[90]

Le PMI dispongono già dei primi strumenti sistematici per tali test interni. Ad esempio, i ricercatori hanno sviluppato un nuovo *framework* software che facilita la pianificazione di esperimenti tra LLM e l'integrazione di LLM in esperimenti con soggetti umani (come dipendenti o clienti).[91] Questo *toolbox* è disponibile gratuitamente e può essere utilizzato, ad esempio, per creare "dilemmi del prigioniero" - un tipico scenario teorico di gioco con molte applicazioni pratiche in economia - di vario tipo, in cui l'interazione di diversi LLM tra loro e le interazioni uomo-macchina possono essere studiate in modo sistematico ed empirico. I risultati mostrano che il comportamento dei modelli linguistici dell'IA può cambiare fortemente e talvolta in modo sorprendente nel corso del tempo o durante interazioni ripetute[92], il che sottolinea l'urgenza di tali test prima che vengano esposti al pubblico. I ricercatori americani hanno anche sviluppato un primo quadro di riferimento per una valutazione complessiva del rischio, che consente di valutare il rischio marginale della messa a disposizione di un modello - cioè il rischio aggiuntivo - rispetto al rischio dei modelli esistenti o della rinuncia totale alle tecnologie di IA.[93]

Quando si conducono test interni, è importante considerare non solo il sistema in sé, ma anche il contesto dell'applicazione.[94] La letteratura mette in guardia, ad esempio, sulle conseguenze di un eccessivo affidamento a modelli errati nella consulenza legale o medica*(automazione e bias di conferma*).[95]Per comprendere meglio il potenziale di tali dipendenze ed errori nel contesto aziendale,

[89]Bitkom (2024), L'AI generativa nell'impresa. Questioni legali relative all'uso dell'intelligenza artificiale generativa nelle imprese, pag. 19.
[90]Si veda lo studio: AI Embraces the "Evil" Side of Online Dating (bsi.ag).
[91]Vedi :GitHub - mrpg/ego : Codice per Engel, Grossmann & Ockenfels.
[92]Engel, Christoph e Grossmann, Max R. P. e Ockenfels, Axel, Integrating Machine Behavior into Human Subject Experiments: A User-Friendly Toolkit and Illustrations (3 gennaio 2024). Documento di discussione sui beni collettivi del MPI, n. 2024/1.
[93]Si veda :On the Societal Impact of Open Foundation Models (stanford.edu).
[94]Dobbe (2022), System Safety and Artificial Intelligence, 2202.09292.pdf (arxiv.org).
[95]O'Neil, C. (2016), Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown; Logg, J. M., Minson, J. A., & Moore, D. A. (2019), Algorithm appreciation: People prefer algorithmic to human judgment.

è opportuno studiare empiricamente i modelli generativi di IA - e il loro utilizzo da parte dei dipendenti - per un periodo più lungo.[96] Uno studio, ad esempio, ha dimostrato che molti modelli possono essere sorprendentemente inclini all'errore in contesti politici, come quando si richiedono informazioni su determinate elezioni.[97] È altrettanto importante prestare attenzione agli errori nella pratica ingegneristica - tra cui la costruzione, la validazione, l'integrazione e la manutenzione - oltre alle possibili carenze nella progettazione teorica.[98] Tali problemi durante l'implementazione pratica sono spesso trascurati nell'attuale dibattito sulla sicurezza dell'IA, anche se è proprio qui che i problemi possono essere affrontati direttamente. Infine, è necessario prendere in considerazione anche l'interazione con le parti interessate esterne all'azienda, come i clienti o le autorità. I sistemi di intelligenza artificiale che interessano aree sensibili e spazi pubblici richiedono una consultazione e una convalida più ampie.

Sulla base di test interni con il proprio sistema di IA, si dovrebbero sviluppare protocolli di interazione e meccanismi di *feedback* per garantire una collaborazione efficace e fluida tra lavoratori e IA.[99] Ciò include la definizione di linee guida chiare su come e quando il sistema di IA debba richiedere l'intervento umano e viceversa. Allo stesso modo, è necessario organizzare corsi di formazione per familiarizzare i dipendenti con gli strumenti di IA generativa e le loro caratteristiche (ad esempio, rispetto a fornitori popolari come OpenAI), dando loro la possibilità di fornire un feedback costruttivo.

2.9 Sostenibilità ed energia: tenere conto dei costi di scalabilità dell'IA

Poiché lo sviluppo dell'IA generativa è caratterizzato da un elevato consumo energetico, questa tecnologia è stata recentemente associata a una crisi ecologica nel settore tecnologico. L'ammissione di Sam Altman, CEO di OpenAI, dell'imminente crisi energetica al World Economic Forum di quest'anno è un buon esempio di questa tendenza ad affrontare la dimensione ecologica dell'IA, sia a livello politico che imprenditoriale.[100] Questa dimensione non si limita all'energia; i sistemi di IA generativa richiedono anche grandi quantità di acqua dolce per il raffreddamento e le principali aziende tecnologiche registrano picchi di consumo significativi per lo sviluppo e l'addestramento dei loro modelli.[101] Queste tendenze sollevano preoccupazioni sulla sostenibilità della rapida crescita dell'IA generativa, in quanto i requisiti di risorse previsti potrebbero raggiungere quelli di un'intera nazione nel prossimo futuro. Di conseguenza, gli esperti chiedono ora lo sviluppo di sistemi di IA più sostenibili, la stesura di relazioni ambientali rigorose e il passaggio a fonti di energia rinnovabili, nonché un'azione legislativa.[102]

Dal punto di vista delle singole aziende, queste considerazioni etiche stanno diventando sempre più importanti quando si utilizza l'IA, sia nei confronti dei dipendenti che dei clienti. Le preoccupazioni

Organizational Behavior and Human Decision Processes, 151, 90-103; Goddard, K., Roudsari, A., & Wyatt, J. C. (2012), Automation bias: a systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, 19(1), 121-127.

[96] Narayanan e Kapoor (2024), AI safety is not a model property (aisnakeoil.com).

[97] Cercate informazioni affidabili sulle elezioni? Non fidatevi dell'intelligenza artificiale (proofnews.org).Raji e Dobbe (2022), Concrete Problems in AI Safety, Revisited, 2401.10899.pdf (arxiv.org).

[98] La vostra organizzazione non è progettata per lavorare con GenAI (hbr.org).

[99] Khalaf (2024), The environmental cost of AI (ft.com).

[100] Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models (arxiv.org).

[101] d41586-024-00478-x.pdf (nature.com).

[102] Ethical Issues in Sustainability-Oriented AI - Anno scientifico 2019: Artificial Intelligence.

etiche si riferiscono ai principi e ai valori morali che guidano il comportamento umano, che ora vengono sempre più incorporati nell'uso dei sistemi di IA al fine di fornire prodotti di IA incentrati sulla sostenibilità.[103] Ad esempio, l'uso di sistemi domestici intelligenti solleva problemi di accesso e utilizzo dei dati, che dovrebbero essere bilanciati rispetto alle opportunità di risparmio energetico. Sebbene esistano molti concetti e raccomandazioni per l'uso etico dei sistemi di IA, come le Linee guida etiche dell'UE per l'IA affidabile e le Raccomandazioni etiche globali sull'IA dell'UNESCO, dell'OCSE e dell'Institute of Electrical and Electronics Engineers, lo sviluppo di normative concrete sta iniziando solo gradualmente.[104] Il concetto di *Value Sensitive Design* (VSD) sottolinea che la tecnologia non è neutrale, ma è impregnata di determinati valori e norme, e mira a integrarli nel processo di progettazione tecnologica.[105] Lo sviluppo dell'IA sulla base della VSD richiede una riflessione critica sui valori e sulle esigenze di tutte le parti interessate e lo sviluppo di sistemi di IA che rispettino tali valori. Per implementare questo concetto a livello di PMI, possiamo pensare, ad esempio, al *toolkit open-source* "AI Fairness 360" di IBM, che offre strumenti per valutare le applicazioni di IA in termini di equità e giustizia.[106] In generale, la letteratura concorda sul fatto che i sistemi di IA responsabili dovrebbero soddisfare standard etici quali equità, spiegabilità e trasparenza[107] e conciliare la sostenibilità ambientale con la redditività e la responsabilità sociale. Le PMI dovrebbero quindi concentrarsi sulla verifica dei dati di addestramento per garantire l'accuratezza degli strumenti di IA e intervenire tempestivamente per evitare pregiudizi e garantire la spiegabilità delle decisioni automatizzate.[108] Il monitoraggio dell'impatto dell'IA sugli obiettivi di sostenibilità e sulla conformità ai futuri standard di settore è essenziale per implementare gli strumenti di IA in modo sostenibile negli appalti e in altre aree di applicazione.

Nella scelta e nell'implementazione delle tecnologie dei modelli linguistici, le PMI dovrebbero quindi tenere pienamente conto del loro impatto sulla sostenibilità e sull'ambiente, sia per motivi etici, di reputazione e di pressione politica[109], sia per i crescenti costi associati alla loro scalabilità. Sia le start-up che le grandi aziende si trovano attualmente ad affrontare costi di implementazione crescenti nel momento in cui passano da un *proof of concept* per pochi utenti a un'implementazione su larga scala della tecnologia dei modelli linguistici.[110] È importante capire che la struttura dei costi del software guidato dall'intelligenza artificiale è molto diversa da quella del software tradizionale.[111] La microarchitettura dei chip e l'architettura del sistema svolgono un ruolo cruciale nella scalabilità della tecnologia dei modelli linguistici. Per questo motivo, l'ottimizzazione dell'infrastruttura dell'IA è essenziale se si vuole utilizzare l'IA generativa in modo sostenibile. Nel prossimo futuro è probabile che vengano introdotti standard che tengano conto non solo della valutazione del consumo energetico diretto e delle emissioni di CO^2 delle tecnologie, ma anche degli aspetti ecologici lungo tutta la catena di fornitura dell'IA.

---

[103]Caso (2023), "Intelligenza artificiale" e sostenibilità sociale. Principi etici per le tecnologie di IA come soluzioni per ridurre la povertà e la disuguaglianza?, Magazin erwachsenenbildung.at 49, pp. 51-60.

[104]Si veda: IA ed etica: la sostenibilità come fattore centrale (suso.academy).

[105]Si veda la panoramica su : IBM Research Trusted AI, AI Fairness 360 (ibm.com).

[106]Si veda : Equità, spiegabilità e trasparenza delle applicazioni dell'IA nel campo della sicurezza: una missione impossibile? - Unione Umanista (humanistische-union.de).

[107]Questa e altre domande si trovano in: AI e appalti sostenibili: la moralità della macchina | Sostenibilità | Haufe.

[108]Si veda :Misurare gli impatti ambientali dell'IA richiede ricerca empirica e standard | TechPolicy.Press

[109]Si veda : Misurare l'impatto ambientale dell'intelligenza artificiale richiede ricerca empirica e standard | TechPolicy.Press.

[110]Si veda: Intelligenza artificiale: Microsoft, Google, Nvidia vincono mentre i costi di calcolo aumentano - Bloomberg.

[111]Questa argomentazione si basa sull'analisi di :Google AI Infrastructure Supremacy: Systems Matter More Than Microarchitecture (semianalysis.com)

Negli ultimi mesi sono stati pubblicati diversi studi empirici che cercano di misurare non solo le emissioni, ma anche altri impatti ambientali e sociali dell'IA generativa, e di sviluppare standard per la rendicontazione degli stessi.[112] Possono essere consultati come punto di riferimento iniziale. Tuttavia, la maggior parte di questa letteratura si occupa solo dei requisiti energetici dell'*addestramento* dell'IA. Per le PMI che incorporano nelle loro attività modelli già addestrati, la ricerca di HuggingFace, che ha quantificato i requisiti energetici dell'*utilizzo dell* 'IA generativa, è particolarmente rilevante.[113] Queste esigenze sono maggiori di quanto si pensi e possono accumularsi rapidamente, a seconda del modello di business e del caso di applicazione. Per le PMI, le tre conclusioni di questa ricerca sono essenziali: 1. le attività di previsione delle categorie sono meno dispendiose in termini di energia rispetto alle attività generative. In altre parole, le applicazioni di IA che consumano più energia e CO^2 sono quelle che generano nuovi contenuti, in particolare la generazione di immagini e (in misura minore) di testi; 2. l'apprendimento dell'IA è ancora di diversi ordini di grandezza. Sebbene l'apprendimento dell'IA sia ancora di diversi ordini di grandezza più dispendioso in termini di energia e di CO^2 rispetto alla singola applicazione dell'IA (nota come inferenza), l'uso diffuso di modelli generativi di IA significa che è possibile raggiungere rapidamente la parità nel consumo di energia per molti modelli comuni; 3. l'uso di modelli generici (come ChatGPT) è più dispendioso in termini di energia per la classificazione del testo e la risposta alle domande rispetto ai modelli specifici.

2.10 Sfruttare le esperienze degli utenti interni e la saggezza delle folle esterne

Infine, le PMI dovrebbero monitorare e valutare regolarmente le prestazioni delle tecnologie linguistiche che implementano. È ampiamente riconosciuto che una riflessione e una valutazione continue aiutano a rispondere agli sviluppi del mercato o della tecnologia, ad adattare le strategie di conseguenza e a identificare le aree di miglioramento. Nel contesto dell'IA generativa e delle PMI qui discusse, sono rilevanti due punti concreti: il monitoraggio dell'esperienza interna dell'utente attraverso la progettazione di modelli *human-in-the-loop* per evitare la "sovra-realizzazione dell'IA"; e l'aggiunta di un'intelligenza collettiva esterna (*"saggezza della folla* ") attraverso i social network, il web e la letteratura prestampata per adattare continuamente la tecnologia linguistica scelta allo stato attuale delle applicazioni e della sicurezza. Che cosa significa esattamente?

A livello organizzativo, l'introduzione di strumenti vocali di IA facili da usare e intuitivi, come quelli di "ChatGPT", può essere problematica a lungo termine, poiché le persone tendono a fare meno sforzi e a prestare meno attenzione man mano che la qualità dell'IA aumenta. Un esperimento sul campo con reclutatori professionisti che esaminavano i CV ha rilevato che coloro che lavoravano con strumenti di IA di qualità inferiore fornivano valutazioni più accurate, in quanto si impegnavano di più e interagivano più efficacemente con l'IA.[114] Inoltre, le persone spesso accettano la decisione raccomandata da un sistema di IA, anche se è sbagliata - un problema che la letteratura descrive come AI*overreliance*, o "fiducia cieca" nell'IA. L'interazione tra uomo e macchina è difficile da valutare ex ante, poiché gli esseri umani non sempre reagiscono razionalmente alle raccomandazioni di un computer.[115] In un esperimento, ad esempio, i partecipanti hanno continuato a seguire i

[112]Si veda ad esempio: Luccioni et al (2022), [2211.02001] Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model (arxiv.org); AI is harming our planet: addressing AI's staggering energy cost (2023 update) (numenta.com).
[113]Luccioni et al (2023), [2311.16863] Power Hungry Processing: Watts Driving the Cost of AI Deployment (arxiv.org).
[114] Dell'Acqua, F. (2022), Falling+Asleep+at+the+Wheel+-+Fabrizio+DellAcqua.pdf (squarespace.com).
[115]La valutazione algoritmica del rischio nelle mani degli esseri umani (iza.org).

consigli dell'algoritmo, deliberatamente mal programmati, anche quando avrebbero dovuto saperlo da tempo.[116] Alcuni ricercatori sperano di ridurre la fiducia cieca nei sistemi di IA costringendoli a spiegare le loro decisioni. Ma i test dimostrano che tali spiegazioni aumentano solo la probabilità che le persone accettino il consiglio dell'IA, che sia corretto o meno.[117] Una soluzione che funziona, almeno a livello sperimentale, è quella di incoraggiare le persone a confrontarsi con queste spiegazioni, anche a livello cognitivo.[118] L'esperto di organizzazioni Gianni Giacomelli ha osservato, a proposito dello sviluppo di modelli *human-in-the-loop* nell'era dell'IA generativa, che "la capacità di sfruttare le nuove possibilità trasformando i nostri processi organizzativi e aziendali e sviluppando le pratiche umane ad essi associate sarà probabilmente importante quanto lavorare sul lato tecnologico dell'IA [traduzione propria]".[119] Secondo una meta-analisi recentemente pubblicata da Microsoft, che raccoglie una sessantina di studi sull'argomento, le misure più importanti per ridurre l *'"eccessiva dipendenza* " dall'IA sono la fornitura di feedback in tempo reale, spiegazioni efficaci per favorire la fiducia e la possibilità per gli utenti stessi di controllare il ritmo e l'uso delle raccomandazioni dell'IA.[120]

Oltre a questo pilastro "interno" per l'apprendimento continuo con e sulla tecnologia linguistica dell'IA, si dovrebbe ricorrere anche a offerte di informazioni "esterne". Soprattutto le PMI, con i loro team e i loro dati talvolta limitati, potrebbero rapidamente scontrarsi con i limiti organizzativi dovuti al rapido sviluppo della tecnologia e a problemi imprevisti. L'utilizzo di un feedback continuo da parte di utenti esterni della tecnologia del modello linguistico implementato o di utenti di sistemi di IA simili è quindi decisivo per la sua efficacia (e accettazione) a lungo termine. Esistono molte audizioni online specializzate che testano quotidianamente le vulnerabilità del LLM (*"Red Teaming* ") e spesso identificano i problemi più velocemente e meglio degli esperti interni. La letteratura accademica non è stata in grado di tenere il passo con questo ritmo per molto tempo; importanti scoperte possono essere trovate in pre-print su ArXiv e vengono discusse su piattaforme di social media come Twitter e Reddit. Le PMI dovrebbero osservare questi discorsi e, se necessario, moderarli attivamente per garantire un'applicazione ottimale ed essere in grado di aggiornare rapidamente i propri sistemi.

3 Conclusioni: sviluppare opzioni strategiche e cogliere le opportunità

Il rapido sviluppo di modelli linguistici di grandi dimensioni come il ChatGPT rappresenta una sfida importante per l'Europa. Data l'applicazione finora insufficiente nell'economia è urgente passare dai "progetti faro" e dall'avversione al rischio a un'implementazione più pragmatica e diffusa della tecnologia dei modelli linguistici.

Nonostante l'intenzione strategicamente astuta dell'UE di garantire la catena del valore dell'IA nel lungo termine con iniziative quali spazi dati, sovvenzioni per fabbriche di chip e supercomputer di IA, le dinamiche della tecnologia dell'IA non consentono ulteriori ritardi nella sua applicazione. L'uso di modelli commerciali e *open source*, in particolare per le piccole e medie imprese (PMI), è essenziale per mantenere la competitività e sfruttare il potenziale dell'IA per automatizzare i compiti ad alta

[116]Biermann, Jan e Horton, John J. e Walter, Johannes, Algorithmic Advice as a Credence Good (2022). ZEW - Centre for European Economic Research Discussion Paper No. 22-071, http://dx.doi.org/10.2139/ssrn.4326911.
[117] L'effetto delle spiegazioni dell'intelligenza artificiale sulle prestazioni dei team complementari (arxiv.org).
[118] Le spiegazioni possono ridurre la sovrarappresentazione dei sistemi di IA nel processo decisionale (arxiv.org).
[119]Giacomelli, G. (2024), Beyond the "human in the loop": reliable AI in enterprise workflows (linkedin.com).
[120]Passi, S. e Vorvoreanu, M. (2022), Overreliance on AI Literature Review (microsoft.com).

intensità di conoscenza e promuovere l'innovazione. In questo contesto, il presente contributo Cep ha descritto dieci fattori che le PMI dovrebbero prendere in considerazione quando implementano la tecnologia linguistica dell'IA. Questi possono essere riassunti come segue:

1. Le PMI devono innanzitutto condurre un'**analisi** approfondita per capire concettualmente come l'IA generativa possa migliorare i loro processi interni ed esterni. Un obiettivo chiaro è essenziale per l'implementazione.

2. **Ridurre le dipendenze strategiche** : la scelta tra servizi basati su *cloud* e installazioni private *on-premise* ha conseguenze dirette su scalabilità, flessibilità e impegni finanziari a lungo termine. L'utilizzo di modelli di base aperti può aiutare a ridurre al minimo le dipendenze strategiche.

3. **Personalizzare i modelli con la messa a punto e la RAG** : le PMI possono ottimizzare e differenziare le loro applicazioni di IA adattando i modelli pre-addestrati a casi d'uso specifici attraverso la messa a punto e la generazione di recupero aumentato.

4. **Sviluppare competenze interne in NLP**: la comprensione del funzionamento dei modelli linguistici e dei loro limiti, nonché lo sviluppo di competenze nella progettazione di *prompt* e nella progettazione online, sono fondamentali per sbloccare il potenziale dell'IA rispetto ai concorrenti.

5. **Adattare la tolleranza all'errore dell'IA**: la tendenza dei modelli di IA ad avere "allucinazioni" e i rischi associati richiedono l'adattamento della tecnologia alla tolleranza all'errore dell'azienda e lo sviluppo di contromisure efficaci.

6. **Integrare gli agenti di IA**: l'aggiunta di agenti di IA ai processi, ai prodotti e ai servizi può migliorare l'efficienza dei flussi di lavoro, ma comporta anche dei rischi che devono essere mitigati attraverso un'attenta valutazione e verifica.

7. **Seguire il quadro normativo**: la legislazione sulla protezione dei dati, sul *copyright* e sull'IA deve essere presa in considerazione quando si implementa l'IA generativa per ridurre al minimo i rischi legali e per poter adempiere agli obblighi di trasparenza. La legislazione sull'IA può essere un'opportunità per le PMI di ottenere maggiore trasparenza sui modelli *black box* sottostanti e sui loro dati di addestramento.

8. Effettuare **test interni** : prima dell'implementazione completa, i modelli linguistici e il loro "carattere" devono essere testati a fondo per garantirne l'accuratezza e l'affidabilità e per individuare tempestivamente gli effetti collaterali indesiderati.

9. **Pensare alla sostenibilità e all'efficienza energetica** : l'impatto ecologico dell'implementazione su larga scala delle tecnologie di IA deve essere preso in considerazione fin dall'inizio per garantire un'implementazione sostenibile ed economicamente vantaggiosa dopo lo *scale-up*.

10. **Utilizzare meccanismi di *feedback*** : la valutazione continua della tecnologia dei modelli linguistici attraverso le esperienze degli utenti e la saggezza delle folle esterne è decisiva per mantenere la tecnologia aggiornata e identificare tempestivamente i vettori di attacco.

L'insieme di questi dieci fattori fornisce alle PMI una base concettuale per sviluppare una strategia interna di IA e sfruttare efficacemente il potenziale nell'area dei modelli linguistici, identificando al contempo i rischi tecnici e le sfide strategiche. La politica europea dovrebbe contribuire a garantire

che la tecnologia dei modelli linguistici possa essere implementata in modo rapido ma sicuro nell'ambiente aziendale nazionale, rafforzando la certezza del diritto (ad esempio, adottando rapidamente linee guida sulla conformità dopo l'adozione della legge europea sull'IA) e fornendo maggiori sussidi e punti di contatto. Allo stesso tempo, le aziende dovrebbero superare il loro scetticismo, sfruttare i progressi metodologici e gli ultimi risultati della ricerca e ridurre la loro dipendenza dai fornitori stranieri per minimizzare i costi di conversione successivi. Questa transizione verso l'applicazione su larga scala dell'IA generativa è essenziale non solo per aumentare l'efficienza e l'innovazione in tutti i settori, ma anche per garantire la resilienza dell'Europa in tempi di instabilità. Se utilizzata su larga scala gestendo attentamente i rischi, la tecnologia dei modelli linguistici ha il potenziale per rafforzare l'Europa nel mercato globale e accelerare la trasformazione verso un ordine economico digitale e sostenibile.

Autore :

Dr. Anselm Küsters, LL.M., Responsabile del Dipartimento digitalizzazione e nuove tecnologie
kuesters@cep.eu

Centrum für Europäische Politik FREIBURG | BERLIN
Kaiser-Joseph-Straße 266 | D-79098 Freiburg

Schiffbauerdamm 40 Räume 4205/4206 | D-10117 Berlin

Tel. + 49 761 38693-0

Traduzione (*dalla versione in lingua francese*):

Dott.ssa Eleonora Poli, *Head of office*
poli@cep.eu

Centro Politiche Europee ROMA
Via A. Brunetti, 60 | I-00186 Roma
Tel. +390636001705
cepitalia@cep.eu

Centrum für Europäische Politik FREIBURG | BERLIN,
Centre de Politique Européenne PARIS
Centro Politiche Europee ROMA
costituiscono il **Centres for European Policy Network** FREIBURG | BERLIN | PARIS | ROMA

Gli istituti della rete CEP sono specializzati nell'analisi e nella valutazione degli atti promossi dalle istituzioni dell'Unione europea nell'ambito delle politiche di loro competenza e nel quadro d'insieme del processo di integrazione. Il lavoro scientifico, riflesso in particolare nelle proprie pubblicazioni, viene portato avanti indipendentemente da qualsiasi interesse di parte e in favore di una Unione europea che rispetti lo stato di diritto ed i principi dell'economia sociale di mercato.